

# Prokaryotic Transcription

## Introduction

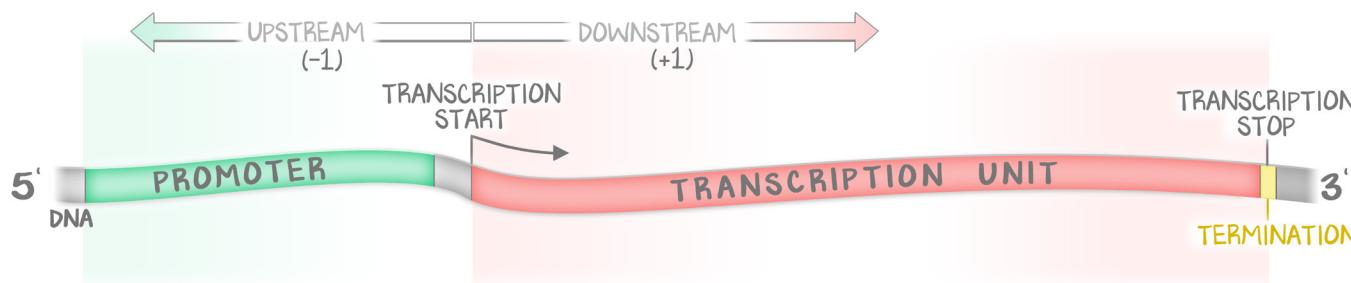
The **template strand** is the one that RNA polymerase uses to build the mRNA. We don't mention the template strand in transcription. The **coding strand** is oriented 5' left, 3' right. The **coding strand is DNA** (has T). The **transcribed strand** is the same as the coding strand, except it's RNA (U instead of T). The mRNA transcribing strand should be oriented 5' left, 3' right. The RNA polymerase starts at the 5' end of its own strand and moves towards its own 3' strand, which is the 5' end of the template strand. Flipped flip discussion in (#7: *Orientation*) isn't repeated here.

Let's focus on what **prokaryotic DNA** looks like and identify all its regions and functions. Then we'll look at what the product of that DNA is, that is, what the **prokaryotic mRNA** looks like. We'll compare the DNA to the RNA and see how they relate. The subsequent lesson (which should ideally be done continuously with this one if time allows) will then look at **eukaryotic DNA** and the differences in making **eukaryotic mRNA**.

This lesson is built foundationally; each section builds on the previous piece. We want to coast in from a bird's-eye view down to the granularity. This differs from the more familiar linear approach. At this lesson's conclusion, you'll have built the foundation for the lesson that follows.

## Prokaryotic DNA

We're talking about what the DNA looks like. We haven't unzipped it, inserted an RNA polymerase, or started to do anything with it. This is what the DNA has in it. The regions will be used by RNA polymerase and instruct it to do a variety of tasks. Because the coding strand and the mRNA are the same thing, we'll discuss only the coding strand and never refer to the template strand. Upstream means "before the RNA polymerase starts." Upstream means to the left of the page. Upstream means the 5' end for the coding strand, and the 5' end for the transcribing strand. 5' left, 3' right for *your DNA*. *Your DNA* means coding DNA.



**Figure 8.1: Bird's-Eye View**

Using this model we'll get more and more detailed, zooming in to each of the sections.

The cistron consists of a promoter region, the transcription unit, and the terminator sequence. The **transcription unit**, all of the DNA from the +1 to termination, is everything the RNA polymerase will transcribe into mRNA. All of the transcription unit will become mRNA; not all of the transcription unit will become protein. The **promoter region** is prior to the start and is therefore listed as negative base pairs. It's where RNA polymerase will bind; it's not transcribed into mRNA. The **termination site** is the site that marks the end of transcription.

## The Promoter Region

Before the transcription unit and upstream of the start signal, the 5' end of the coding DNA strand (which means the 5' end of the transcribing mRNA strand), there's an area called a **promoter region**. The promoter region consists of two **consensus sequences** (also referred to as “**boxes**”). There's one at -35 (35 base pairs upstream of the 5' end of the start signal on the coding DNA) and one at -10 (10 base pairs upstream of the 5' end of the start signal on the coding DNA). The one at -10 is called the **TATA box**.

There are three pieces required to let the RNA polymerase bind. The two consensus sequences are required first to **tell RNA polymerase where to go**. This would require only one region simply to bind. But the second consensus sequence is required to **orient RNA polymerase**. Binding of RNA polymerase to the promoter region also requires the presence of a  **$\sigma$  (sigma) factor**.



**Figure 8.2: Promoter Region in Detail**

Engaging the promoter region in more detail, we see the presence of the TATA box and other consensus sequences.

## The Transcription Unit

The **transcription unit** is all of the DNA that'll become mRNA. It consists of the coding region, the 5' untranslated region (5'UTR), and the 3' untranslated region (3'UTR). For polycistrons (at the end) there are regions of untranslated DNA between coding regions.

A three-nucleotide-long sequence that pairs with a tRNA's anti-codon is called a **codon**. Codons are more important in translation, where we discuss them in detail.

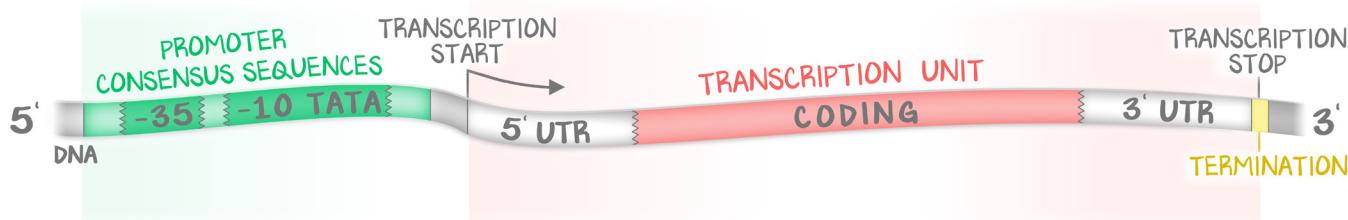
The **coding region** is the DNA that'll be transcribed and then used for translation—DNA that's made into mRNA, eventually to be translated into amino acid. The coding region will become the active mRNA that interacts with ribosomes and pairs with tRNA. The coding region always starts with an **initiation codon**, the “start” signal. The coding region always ends with a “stop” signal. “Start” and “stop” codons signal the start and stop of translation.

The **untranslated regions** are un-translated. They're **transcribed**—they'll exist in the mRNA sequence—but they won't interact with tRNA, and that section of code won't yield an amino acid or be part of a protein. There's a **5' UTR** (5' on the coding DNA strand, 5' on the transcribing RNA strand), and a **3' UTR** (3' on the coding DNA strand, 3' on the transcribing RNA strand).

The **transcription termination signal** comes in one of two mechanisms.  **$\rho$  (rho)** is a protein that can bind to newly formed mRNA, chases down the RNA polymerase, and when it catches up, displaces the RNA polymerase from the chain.  $\rho$ -dependent termination involves  $\rho$ .  **$\rho$ -independent** termination (bolded and emphasized because this is the one to pay attention to) involves a structural change in the transcribing RNA sequence that causes the mRNA strand and the RNA polymerase to dissociate.

from the DNA. The termination of transcription is caused by a structural change in the mRNA being transcribed—transcription causes the end of transcription.

Transcription begins with the first nucleotide of the transcription unit, designated **+1**, the first nucleotide of the 5' UTR. Transcription ends at the **termination site**, which is a variably sized sequence of the 3'UTR.



**Figure 8.3: Transcription Unit in Detail**

The transcription unit is made of a 5' untranslated region, a 3' untranslated region, and a coding region.

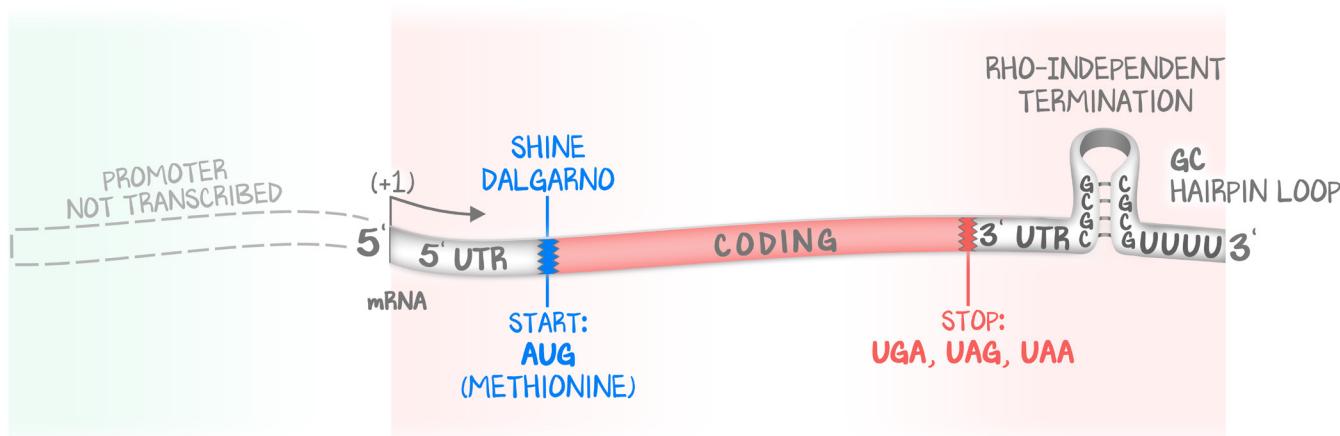
## What Prokaryotic mRNA Looks Like

RNA polymerase binds to DNA that is helical and zipped. RNA polymerase unzips it, reads it, and makes the mRNA transcript. The RNA polymerase binds at the promoter and begins transcription with nucleotide +1.

The **promoter region** is **not copied**, so that region is **absent in mRNA**.

The **transcription unit** is entirely transcribed, so all of the DNA pieces have an mRNA correlate.

The **5' UTR region** of the coding DNA strand codes for the 5' UTR region of mRNA. That 5' UTR region of mRNA carries the **Shine-Dalgarno sequence**. From the +1 site to the Shine-Dalgarno sequence in the 5' UTR region of mRNA is simply a strip of nucleotides that allows the ribosomal RNA (during translation) to grab hold of the strand and start reading. The Shine-Dalgarno sequence says, “*as soon as this sequence is over, the first codon will appear. This is the time you should start adding amino acids.*” The **start codon**, appearing just after the Shine-Dalgarno sequence, is the beginning of the coding region.

**Figure 8.4: mRNA in Detail**

The promoter region is not transcribed, so is absent, and the 5' untranslated region has a Shine-Dalgarno sequence before the start of the coding sequence. The start of the coding region (which will be translated) always begins with an AUG and ends with UAA, UAG, or UGA. Also demonstrated is the p-Independent transcription termination sequence, the GC-hairpin loop.

The **coding region** carries all of the mRNA that'll eventually be translated into protein. The coding region is the active mRNA that interacts with ribosomes and pairs with tRNA. The coding region always starts with an **initiation codon**, the “start” signal **AUG**, which codes for a **methionine**. There’s one and only one initiation codon. Every “start” can be only AUG and every “start” can be only methionine. And since mRNA turns into protein, every protein must begin with methionine. The coding region **always ends with a “stop” signal**. “Stop” is one of three codons. The stop codon sequences are worth memorizing. Memorize the pattern, the codons, or both. The stop sequence: **“Always start U, never include C, repeat only A”** or **“UAG, UGA, UAA”**.

The **3'UTR** of mRNA after the coding region has extensive uncoded mRNA. This is the area where we find the two p-independent mRNA structural features. The first is the **hairpin loop** formed by a series of GC’s. These GC’s are coded in sequence—G-C-G-C-G-C—repeated. The loose, single-stranded mRNA has so many GC repeats, and the attraction between G and C (three hydrogen bonds) is so strong, that this sequence acts a magnet to itself and forms base pair bonds. The single-stranded mRNA snaps together so that it appears as though it’s double-stranded. The turn makes a loop that ejects the RNA polymerase.

The **second** is a series of **UUUUUU**, which came from the termination sequence of the coding DNA TTTTTT. The combination of the hairpin loop and UUUUUU marks the end of the mRNA; it’s the terminator site for RNA polymerase.

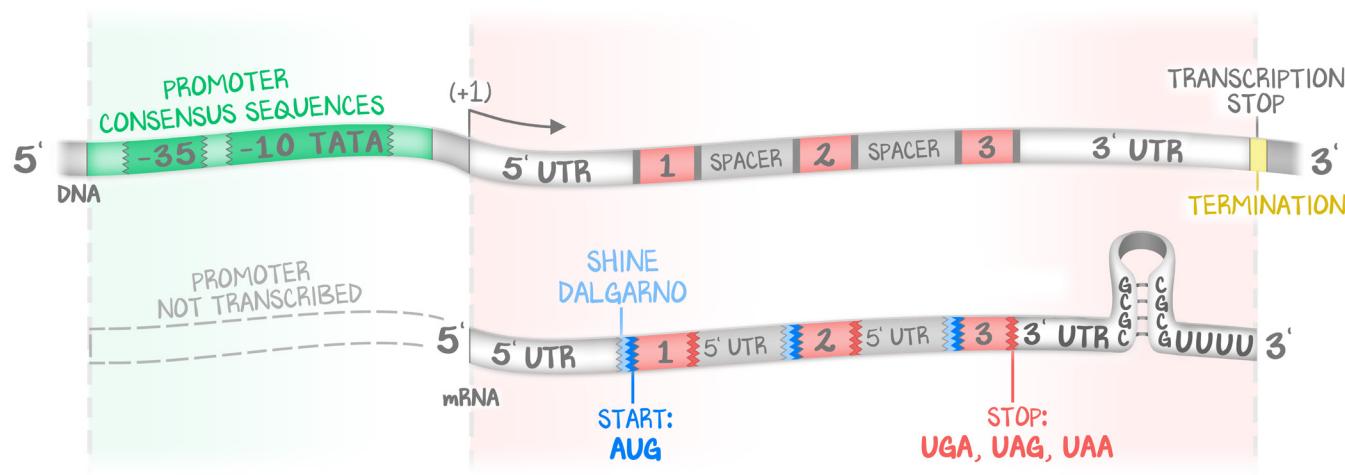
## Cistrons

When **one promoter region** codes for **one gene**, one protein, it’s called **monocistronic**.

When **one promoter region** codes for **multiple genes** it’s called **polycystronic**. Because all genes have only **one promoter region**, all of the genes in a polycistronic sequence will be **transcribed equally**. Every time the RNA polymerase binds to the promoter region, each of the genes will be transcribed. There’ll be a **Shine-Dalgarno sequence for each gene** and a **stop codon** for each gene, but there’ll only be **one 5'UTR** and **one 3'UTR** with a hairpin loop and UUUUUU.

Since there's a Shine-Dalgarno sequence for each gene, a ribosome-making protein from that gene will start there, and it'll stop at the stop codon. There are strips of mRNA that aren't used, which can be seen as relative 3'UTR's (since it's not a 5'UTR loop and does have a Shine-Dalgarno).

And because prokaryotic transcription and translation occur in the cytoplasm, **transcription and translation occur simultaneously**. Ribosomes bind to each Shine-Dalgarno region separately, such that while one promoter transcribes one mRNA strand at the same time, each gene, each Shine-Dalgarno sequence, each start-to-stop will have a discrete ribosome. If there are three genes, three Shine-Dalgarno sequences, then three ribosomes can be working on three separate genes at the same time; one ribosome will not continue onto the next gene.



**Figure 8.5: Polycistrons**

Polycistronic mRNA contains multiple coding sequences separated by untranslated regions that are all under the regulation of one promoter region.